11-1-2019

Emergent functional behaviour of humic substances perceived as complex labile aggregates of small organic molecules and oligomers

Robert L. Cook

Louisiana State Univ, Dept Chem, rlcook@lsu.edu

CSIRO PUBLISHING

Environ. Chem. **2019**, *16*, 505–516
https://doi.org/10.1071/EN19095

Research Paper

Emergent functional behaviour of humic substances perceived as complex labile aggregates of small organic molecules and oligomers

Elena A. Vialykh,[A,G] *Dennis R. Salahub,*[B,C] *Gopal Achari,*[D] *Robert L. Cook*[E] *and Cooper H. Langford*[A,F]

[A]Department of Chemistry, University of Calgary, 2500 University Dr NW, Calgary, AB, T2N 1N4, Canada.
[B]Department of Chemistry, CMS — Centre for Molecular Simulation, IQST — Institute for Quantum Science and Technology and Quantum Alberta, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 1N4, Canada.
[C]College of Chemistry and Chemical Engineering, Henan University of Technology, 100 Lian Hua Street, High-Tech Development Zone, Zhengzhou 450001, China.
[D]Schulich School of Engineering, University of Calgary, 2500 University Dr NW, Calgary, AB, T2N 1N4, Canada.
[E]Department of Chemistry, Louisiana State University, 307 Choppin Hall, Baton Rouge, LA 70803, USA.
[F]Deceased.
[G]Corresponding author. Email: elena.vialykh@ucalgary.ca

Environmental context. The correlation of physicochemical characteristics of humic substances with their function is crucial to our understanding of how environmental pollutants interact with humic substances. We have developed an approach that models emergent functions of fulvic and humic acids depending on sample characteristics. The results will be useful for predicting the sequestration of organic contaminants in soil under various conditions.

Abstract. The structural organisation of humic substances (HS) has been a central question of earth sciences for several decades. The latest experimental results have led to the recognition of HS as complex mixtures of small molecules and oligomers. We investigate the correlation between the chemical composition of HS, perceived as labile aggregates, and the emergent functions. Computational modelling was used to help to understand the processes and mechanisms on the molecular scale that occur in different fractions of the HS, fulvic acids (FA) and humic acids (HA), as they interact with metal ions and organic pollutants. The importance of non-covalent interactions in the emergent functions of HS is highlighted. H-bonding, hydrophilic/hydrophobic surface areas and π-stacking interactions play a significant role in aggregation processes as well as in the sorption of environmental pollutants. In a highly hydrophilic system with small molecules (the SRFA-22 model), H-bonding is the main force that drives the aggregation process. However, in a highly aromatic and hydrophobic model with larger molecular fragments (SRHA-6), hydrophobic and π-stacking interactions dominate in the aggregation process. The chemical properties of contaminants significantly affect their mechanisms of sorption by HS. The interaction of a polar pollutant, phenol, with HS occurs through H-bonding, whereas non-polar benzene interacts through hydrophobic and π-stacking interactions. The non-polar pollutant results in a much stronger sorption by HS and causes an additional structural rearrangement of the aggregates, which make it more stable in the environment.

Received 25 March 2019, accepted 16 May 2019, published online 11 June 2019

Introduction

The organic components of soil form a complex mixture of thousands of different organic molecules (Drosos et al. 2017; Hertkorn et al. 2008; Reemtsma 2009). The two best-characterised fractions of humic substances (HS) (first identified by base extraction in the 19th century) are denoted as fulvic acids (FA) and humic acids (HA). Soil organic matter is responsible for several major soil and water properties, which include soil water retention and sequestration of xenobiotic toxicants (Perminova and Hatfield 2005). A recent paper has suggested that HS, with the capacity to sequester hydrophilics, hydrophobics and metal ions, function as 'universal protectors of the ecosystem' (Tan 2014).

For these reasons, studies of HS have been conducted over many decades. Since HS are mainly derived from biological materials, an attractive approach has been to regard them as covalent polymers analogous to their precursors. However, the development of newer techniques to examine samples (Cook

and Langford 1998; Nebbioso and Piccolo 2012; Underdown et al. 1981; Wershaw 1986) have challenged the covalent polymer models. The final critical step was the analysis with high resolution, Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry that revealed over 4000 distinct molecular formulae in one of the simplest HS samples, Suwannee River fulvic acid (Stenson et al. 2003).

Contemporary understanding (Piccolo 2002) recognises HS as complex mixtures of small molecules and oligomers that can form labile aggregates. Continuing research has yielded an expanded and detailed appreciation of the composition of these mixtures (Reemtsma 2009). The main difficulty encountered in studies relates to the complexity of HS (Orsi 2014). HS are highly polydisperse in composition, and one sample may contain thousands of different individual molecular fragments (Hertkorn et al. 2008) that vary in molecular size. The observed properties of HS depend on the labile aggregates that form under weak intermolecular forces.

Molecular dynamics (MD) simulations have offered insight into several HS phenomena (Orsi 2014; Schaumann and Thiele-Bruhn 2011). In most of the computational work performed recently (Alvarez-Puebla et al. 2006; Petrov et al. 2017; Sutton et al. 2005), the authors recognise explicitly the concept of the aggregate conformation of HS. However, they mainly addressed the modelling of certain chemical properties by using a single structure to represent the overall properties of the mixture (Alvarez-Puebla et al. 2006). Only a few publications have attempted to probe the structure and dynamics of HS samples as well as to study their physicochemical properties by implementing a supramolecular concept, where the behaviour of the system is determined by weak forces (Petrov et al. 2017). As a consequence, an appreciation of the mechanisms behind the functions has not advanced synchronously with the structural understanding. Analysis of the functions remains uncertain and speculative, since both the molecular structure and the supramolecular arrangement of the various molecules in the complex control these functions (Schaumann and Thiele-Bruhn 2011).

This work proposes an approach for developing the basic mechanistic insight to support the understanding of the important functions of HS. An aggregation dynamics approach combines two elements. First, the current structural understanding (e.g. from FT-ICR MS) allows the selection of representative individual structures to use in modelling and, second, treats the interaction of smaller and medium-size molecules to form labile aggregates as important for the representations of functions. This approach is theoretically based on ideas of agent-based modelling and dynamic combinatorial chemistry (Corbett et al. 2006; Huang et al. 2005).

The core idea of agent-based modelling is that individual agents, in this case, smaller molecules, can have manageably small sets of specified properties that govern their interactions, and that the complex and emergent behaviour of a system can arise from the specified interactions of a set of agents interacting over time. Moreover, these interactions and the influenced entities will feed back on the composition, structure and turnover of HS (Bar-Yam 2002; Corbett et al. 2006; Schaumann and Thiele-Bruhn 2011).

Finally, the ideas of agent-based modelling and recognition of HS as labile aggregates of small molecules are implemented with computational modelling, an approach that has had success in illuminating the functions of a variety of complex biological systems. Thus, computational chemistry could be a highly versatile tool that helps to understand processes, and the involved mechanisms, sites and interactions on the molecular scale within a 3D structural network. The goal of this work is the creation of reliable computational models that mirror the properties of larger labile supramolecular assemblies (Piccolo 2002; Sutton and Sposito 2005), and the use of the model to study the mechanisms of HS interactions with representative 'pollutants', xenobiotic organics or metal ions. The modelling in this work is of an exploratory nature and is the first step in the elucidation of some mechanisms of the characteristic behaviours of HS and how emergent properties can arise. Therefore, the theoretical background is described more explicitly and underpinned with computational modelling results without attempting to obtain statistical convergence of the MD simulations and calculations of free energy. The reference 'real' sample is the International Humic Substances Society (IHSS) Suwannee River fulvic acid (SRFA) and humic acid (SRHA) from the reference collation of the IHSS, as these samples are, in all likelihood, the most studied HS (IHSS 2019).

Experimental

Computational design: theory and selection of force field

The principal idea of an agent-based model is that for a complex system consisting of individual agents, the behaviour of the system is determined by the specific properties and interactions assigned to each agent. Interactions include those with other members of the system and with the environment (Cabaniss et al. 2005). Computational chemistry methods assign properties to the agents by assigning classical 'force fields' to the molecules. The supramolecular model focuses on the labile interactions between molecular units. The model must include van der Waals forces, hydrogen bonding, charge-transfer interactions, steric factors, π-stacking, hydrophobic and hydrophilic interactions, water solvation, and metal-ion complexation. Additionally, one can anticipate proton transfer reactions and esterification. To include these interactions, a newer force field, ReaxFF (van Duin et al. 2001), was suggested. ReaxFF allows chemical reactions to be modelled by determining the bond orders from interatomic distances that are updated at every step in the MD calculations. Therefore, ReaxFF is able to simulate bond formation and breakup during modelling (van Duin et al. 2001). In this work, the ReaxFF force field parameters for modelling proton transfer reactions in organic molecules in aqueous environments as well as other aspects of biologically important phenomena, which includes metal complexation, were used (Rahaman et al. 2011).

All members of a system are in labile equilibrium with respect to the weak intermolecular forces considered. They may undergo conformational changes, aggregation and chemical transformations with a change of environmental conditions. Therefore, the composition of a system is determined by the thermodynamic stability of each library member under environmental conditions (Corbett et al. 2006). Control implies that changing the experimental conditions can induce changes in the distribution of species, which causes the system to respond to external influences (Corbett et al. 2006).

Choice of molecular systems

In the conventional approach for modelling an HS mechanism, various data are used to create a single molecular unit that would most adequately represent the HS functionalities to be explored (Jansen et al. 1996; Sein et al. 1999). This ignores the high polydispersity revealed in the structure as well as the evidence

for labile supramolecular assemblies of molecular fragments with independent conformations. Consequently, emergent features cannot be accommodated. The generation of various models to reflect the diversity in composition and conformations of the constituting molecules has only been performed in a few recent works (Petrov et al. 2017; Sündermann et al. 2015).

This work studies three different systems that are constructed from 6 or 22 molecular units and vary in composition and molecular size (Table S1, Supplementary Material). The guidelines for the choice of molecules were adequate fits to elemental composition, nuclear magnetic resonance (NMR)-derived functional group distribution, and aromatic to aliphatic and 'carbohydrate like' ratios. The molecular mass distribution is close to the values from FT-ICR MS of 100–700 Da for SRFA (Reemtsma et al. 2008; Remucal et al. 2012), with a bias towards formulae reported from the high-resolution MS studies.

The first system, SRFA-22, was built by using molecular fragments identified by FT-ICR MS analysis (Remucal et al. 2012) with the addition of 3 carbohydrate moieties to generate a better representation of the NMR spectra. Out of the 172 identified fragments, the 10 most abundant fragments were chosen in a way that the overall chemical characteristics fitted to the average parameters of SRFA, that is, the reference sample provided by IHSS (Table S2, Supplementary Material). The molecular weights of fragments varied between 110 and 400 Da. Similar to Sündermann et al. (2015), the following criteria of the system were used to build an overall SRFA-22 model: C, N and O content and functional group compostion of SRFA. Though, in our work, the fragments identified by FT-ICR MS (Remucal et al. 2012) were used to generate the SRFA-22 model.

Molecular fragments for the second model, SRFA-6 (Table S1, Supplementary Material), were selected from the structures proposed by Leenheer et al. (1994). These hypothetical structures were designed by integrating the state-of-the-art knowledge from the analytical data on SRFA (Text S1, Supplementary Materials). The main difference from the previous system was in the size of the molecular fragments, though the molecular weights of the systems overall were similar. The third system, SRHA-6, was created by modifying the molecular fragments of the second system so that overall, it fitted to the average SRHA characteristics. However, the molecular fragment sizes remained similar to the SRFA-6 model and varied between 660 and 840 Da (Table S1, Supplementary Material).

Computer simulation method

The implementation of the dynamic combinatorial chemistry approach for HS modelling, while limiting the initial models to 6 or 22 molecular units, raises another issue related to the choice of molecular simulation methods. Molecular modelling methods that allow for the simulation of systems containing thousands of atoms, such as AMBER (Wang et al. 2004) and CHARMM (Vanommeslaeghe et al. 2010), are inherently incapable of describing chemical reactions. To overcome this, the ReaxFF force field was used (van Duin et al. 2001). ReaxFF force field parameters are optimised to reproduce a suitable reference organic dataset, derived mainly through quantum chemical calculations. Being empirical in nature, ReaxFF allows MD simulations of large reactive chemical systems (1000s of atoms) and yet retains an accuracy close to that of quantum chemical calculations.

Since the HS of focus here consist of organic molecules in an aqueous environment, we used the ReaxFF potential developed to simulate biomolecules (Monti et al. 2013). Nevertheless, we tested ReaxFF for its ability to simulate molecular interactions that are important for complex organic mixtures and that occur in HS, which include (1) H-bonding; (2) hydrophobic interactions; (3) π-stacking; (4) charge transfer complexation; (5) labile metal-ion complexation; and (6) chemical reactions between molecular fragments. We used ReaxFF to study the functional behaviour of three different systems with organic contaminants, phenol, toluene and benzene, in a vacuum and aqueous phase, as well as the complexation with Cu^{2+} in a vacuum.

The initial configurations of the input structures were generated with the *PackMol* program (Martínez et al. 2009). The initial packing of the models to perform MD simulations in a vacuum was with the following condition: the distance between the closest atoms of two different molecules could not be less than 4 Å, to ensure the minimum initial interactions between molecules. The SRFA-22 model contained 22 molecular fragments, whereas SRFA-6 and SRHA-6 contained six fragments each. The side of the periodic box (cube) was 50 Å for models simulated in a vacuum. All the systems were geometry optimised before MD simulations. Geometry optimisations were deemed to have converged when the root mean square of forces fell below 2.1 kJ mol^{-1} $Å^{-1}$. Molecular dynamics simulations were carried out in the NVT (constant number of particles, volume and temperature) or NPT (constant number of particles, pressure and temperature) ensembles, in a cubic simulation box with periodic boundary conditions. The simulations were carried out at temperatures in the range 300–600 K. All the calculations used an MD timestep of 0.25 fs. During simulations, the system temperature was controlled by a Berendsen (for simulations carried out in vacuum) or Nose–Hoover (for simulations carried out in aqueous solution) thermostat with a damping constant of 0.1 ps.

The dissolution MD simulations were carried out as follows. *PackMol* was used to place the aggregates formed in a vacuum in an aqueous environment, by adding 1000 water molecules to the system, with approximate concentrations of the organic part of 22–25 %. The side of the periodic box (cube) was 60–65 Å. Each generated system was pre-equilibrated for 0.1 ns under NVT conditions with $T = 300$ K, followed by simulations performed under NPT conditions with $P = 1$ atm, and $T = 300$ K for another 0.2 ns (or until the size of the periodic box stabilised, i.e. variation in one dimension less than 1 %) to obtain the density of the systems under standard conditions; and final pre-equilibration of the MD simulations was carried out under NVT conditions for 0.5 ns with the size of the periodic box obtained from the previous NPT step. The following equilibrating NVT-MD simulations were carried out for 1.5 ns, where after 1 ns of MD simulations at 300 K, the temperature was raised to 400 K for 0.25 ns to accelerate the dissolution process, and then the system was cooled to 300 K and allowed to equilibrate for 0.25 ns. The hydrophobic/philic surfaces areas, total potential energy and number of hydrogen bonds of the systems were evaluated at the beginning of the equilibrated MD simulations, at the end of modelling at 400 K and at the end of the MD simulations after cooling the system. All simulations were conducted at least in triplicate with different initial packing and starting velocities.

The *VMD* tool for hydrogen bond analysis (Humphrey et al. 1996) was used to estimate the number of hydrogen bonds in the mixture. A hydrogen bond was considered to be formed between an atom with a hydrogen (the donor, D) and another, mainly O, atom (the acceptor, A) provided that the distance D-A was less

than the cut-off distance of 3.0 Å, and the angle D-H-A was less than the cut-off angle of 20° (Gumbart 2007). *Chemsketch* software (Hunter 1997) was used to draw the 2D structures of molecules. Hydrophilic/phobic surfaces were determined according to the procedure described by Aristilde and Sposito (2010) by using the *Maestro 10.2* software (Schrodinger Software 2015). The average values of the surface areas of each system were calculated.

To calculate the number of hydrogen bonds, molecular and atomic charges, the total potential energy and the hydrophobic/hydrophilic surface areas for each system, 20 snapshots from the last 0.10 ns of the MD trajectories were selected at equal intervals. The average values of each characteristic were calculated. A description of the ReaxFF force field is presented in the Supplementary Material (Text S2).

Results and discussion

Simulation of the interactions peculiar to HS aggregation: proof of concept

A computational model must be able to accommodate conformational flexibility as well as the interaction mechanisms that have been proposed to account for labile supramolecular structures, which include H-bonding, hydrophobic interactions, π-stacking, charge-transfer complexation, proton transfer and labile metal complexation. The first simulations using the ReaxFF package yielded examples of each of these phenomena. Although performing simulations in a vacuum may not seem particularly realistic, the vacuum simulations may offer some insight into the interaction mechanisms that predominate in the dry state. An important aspect of HS chemistry is the difference between the behaviour of wet versus dry systems (Litvina et al. 2003). Neglect of solvation is a justifiable limit to obtain examples of the interactions involved in HS. In the subsequent MD simulations, water was added to the aggregates obtained in a vacuum to model the dissolution process of dry organic matter.

H-bonding, aggregation and conformational flexibility

First, we modelled the behaviour of the mixtures in a vacuum at 300 K. During the simulations, all three models tended to form chaotic aggregates with hydrogen bonds between molecular fragments with a decrease in the number of hydrogen bonds in the order SRFA-22 > SRFA-6 > SRHA-6 (Table S3, Supplementary Material). The higher number of hydrogen bonds found in SRFA-22 (21 ± 2 versus 17 ± 2 in SRFA-6), despite its lower number of hydrogen-bond-forming groups (48 in SRFA-22 versus 55 in SRFA-6), is explained by the fact that this model is composed of smaller molecules and hence is more flexible. This flexibility allows it to arrange itself to maximise the number of hydrogen bonds. Stated another way, an emergent property of a system composed of smaller molecular moieties allows the functional groups on these moieties more degrees of freedom compared with their counterparts with large molecular entities. Therefore, the additional flexibility results in a higher possibility of the functional groups to be in an arrangement that is conducive to hydrogen bonding.

The value of the total potential energy was obtained to examine the energetic properties. Although the parameters in ReaxFF are not referenced to a unique molecular state (i.e. infinite separation of interacting species), it is important to note that the energy values it computes can be compared only relatively, within a given set of simulations. Thus, the changes of total potential energy (ΔE_{pot}) computed from MD simulations were used to assess the relative thermodynamic stability of the obtained configurations of the formed labile aggregates and could be considered as a property that reflects the emergent functions of the complex organic mixtures, which includes aggregation. Moreover, the potential energy is one of the parameters directly provided through MD simulations and can be used for a fast evaluation of the changes occurring in the model. The interaction energy (difference between the sum of total potential energies of individual fragments and average potential energy of the model during the last 0.1 ns of MD simulations) for all models was between −493 and −949 kJ mol^{-1} (Table S3, Supplementary Material). Comparing SRHA-6 and SRFA-6, it can be seen that SRFA-6 has a greater potential energy drop of −949 kJ mol^{-1} than that of SRHA of −494 kJ mol^{-1}. This difference can be explained by the large number of hydrogen bonds associated with SFRA-6 versus SRHA-6 after MD simulations. A comparison of the potential energy drop of SRFA-22, −790 kJ mol^{-1}, with that observed for SRFA-6 shows that hydrogen bonding does not fully explain the observed potential energy drop. There are also other interactions responsible for the energy decrease, which will be discussed below.

An attempt to determine the global minimum for the aggregates (MD simulations carried out for additional 1 ns at 300 K) revealed that there were at least 31 structural conformers within an energy range of 84 kJ mol^{-1} of the optimal structure for all three models. Similarly, in other works (Bruccoleri et al. 2001; Schulten et al. 2001; Schulten and Leinweber 2000), attempts to find the global minima of HS molecular fragments or aggregates demonstrated the formation of flexible networks with voids, with several assemblies with very close energies. Thus, for HS, and natural organic matter (NOM) as a whole, there will not be a highly preferred assemblage under environmental conditions and, more importantly, there will be several different assemblies present at thermal equilibrium with concentrations of the same order of magnitude as the optimum assembly at any point in time. Thus, the dynamics and flexibility of the HS appear to be decisive properties, which emerge as a suite of molecular assemblies of HS, and NOM in general, in the environment. The extraordinary flexibility and structural variation in the model structures underlines the importance of inter- and intramolecular non-covalent interactions and helps us to understand how molecular assemblies have, as a collective and working in concert, such a wide and diverse set of highly useful and valuable environmental functions.

Hydrophobic, π-stacking and proton transfer interactions

In this section, we discuss the non-hydrogen bonding interactions, which were foreshown in the section above. To accelerate MD simulations and observe possible additional (slower process) interactions between organic molecules, we performed MD simulations, in a vacuum, of the systems described above at a temperature of 600 K followed by cooling to 300 K and equilibrated for 1 ns according to the temperature profile (Fig. S1, Supplementary Material) (Ashraf et al. 2017; Bruccoleri and Karplus 1990).

These higher temperature simulations resulted in several SRFA-22 assemblies in which the number of hydrogen bonds remained unchanged at 21 ± 2; however, there was a decrease in energy of 158 kJ mol^{-1}. This decrease in energy was accompanied by the formation of π-stacking structures, which involved no more than two aromatic-ring-containing molecular fragments per structure (Fig. S2, Supplementary Material). The

Table 1. Change in characteristics of organic models after MD simulations with raising temperature to 600 K

	Change in hydrogen bonds[A]	Change in potential energy, $\Delta E_{pot} = E_{pot,600K} - E_{pot,300K}$ [A] (kJ mol^{-1})	Change in hydrophobic/ hydrophilic surface[A] (Å^2)
SRFA-22	0	−158	190/−147
SRFA-6	−1	−766	−418/−209
SRHA-6	−2	−728	−303/−68

[A]The value calculated as the difference between average value obtained for the system after MD simulation at 300 K and average value obtained for the system after MD simulations with temperature increased to 600 K and following cooling down to 300 K.

associated rearrangements resulted in an increase of 31 % in the hydrophobic surface area and a decrease of 23 % in the hydrophilic surface area, which translated to the surface of the assembly becoming more hydrophobic. This is consistent with what one would expect, as the simulations were carried out in a vacuum, a non-polar environment, which resulted in less charged (more hydrophobic) sites accumulating outside and driving interacting polar groups (more hydrophilic) to the interior. Overall, the temperature increase led to the formation of a system with decreased potential energy, which emerged as a result of structural rearrangements resulting from several interactions, including π-stacking and hydrophobic/philic interactions that take into account van der Waals forces and electrostatic interactions (Schrodinger Software 2015).

Quite a different, but significant, result related to the temperature increase was obtained for SRFA-6 compared with SRFA-22. Raising the temperature during simulations promoted the formation of a more organised complex structure with π-stacking interactions between aromatic parts of 3 or 4 molecular fragments (Fig. S2, Supplementary Material). The distance between the aromatic rings, which were involved in π-stacking, varied between 3.4 and 4.5 Å and was, as expected, affected by conformational hindrance caused by other parts of the molecules. The number of hydrogen bonds in SRFA-6 decreased by 1. Additionally, the total surface area of the assembly decreased with the hydrophobic and hydrophilic surface areas decreasing respectively by 31 % and 23 %, and the overall potential energy decreasing by 766 kJ mol^{-1} (Table 1). The major driver for these changes in the assembly appeared to be the formation of π-clusters and hydrophobic centres; however, other non-covalent interactions could be at play as well. In the more aromatic SRHA-6 model, π-stacking interactions between the aromatic parts of up to four molecular fragments were observed. This resulted in the potential energy of the model decreasing by 728 kJ mol^{-1} with a more compact molecular assembly and a decrease in the hydrophobic and hydrophilic surfaces areas of 25 % and 9 % respectively. Additionally, beyond hydrophobic/hydrophilic, van der Waals and electrostatic, and the above discussed π-stacking interactions, proton transfer occurred (Fig. S3, Supplementary Material) in each of the models investigated. Thus, a resonance induced intramolecular keto–enol proton transfer interaction was observed in the SRFA-22 and SRHA-6 models.

Charge-transfer interactions and metal-ion binding

Introducing metal ions to the complex organic mixtures with appropriate metal binding functional groups, such as those found in HS, results in metal-ligand complexation, which in turn leads to more compact assembly and aggregation (Cabaniss 2011; Xue and Sigg 1999). In this work, we used ReaxFF to evaluate the possibility of a Cu ion, with a formal charge of 2+, which interacts with different functional groups in our SRFA and SRHA models, by modelling, in a vacuum at 300 K for 2 ns. Cu^{2+} initially was bound to two deprotonated carboxylate groups, to maintain an overall system charge of 0 (Fig. S4, Supplementary Material). For all systems, SRFA-22, SRFA-6 and SRFA-6, two additional Cu-O bonds were formed, which led to the formation of a larger-sized complex that was stable over MD simulations (Fig. S4, Supplementary Material). This result was expected, since Cu^{2+} forms tetrahedral complexes with organic ligands (Bertoli et al. 2016), and the preferred functional group for complexing Cu^{2+} is the carboxylate group. For the two FA systems, carboxylate groups associated with alkane fragments and a phthalate group were also involved in Cu^{2+} complexation, while for the HA system, Cu^{2+} complexation involved the carboxyl group of a benzoate fragment. The formation of coordination bonds between copper and ester groups in all three systems were also observed. This involvement of ester groups might arise from all of the carboxyl groups in the models, except the two bound to Cu^{2+}, being protonated, and the MD simulations being performed in a vacuum. These findings agree with those observed for Cu^{2+} binding to (i) FA as monitored by NMR, in which it was shown that carboxylate groups associated with alkane fragments were the major group involved in Cu binding, especially at lower Cu concentrations (Cook and Langford 1999*a*, 1999*b*) and (ii) soil as monitored by Fourier-transform infrared spectroscopy, in which Cu contamination resulted in an increase of asymmetric and symmetric COO^- stretching and the reduction of the O–H stretching of H-bonded, C=O stretching and C–O stretching functional groups that participate in Cu binding at low pH (Alvarez-Puebla et al. 2004). This alignment of simulation with experiment supports the simulation approach used here and provides confidence in the results of the simulations carried out in this study.

Although there were similarities in behaviour with and without Cu^{2+} ion, such as the formation of the same number of H-bonds (21 ± 2, 17 ± 2 and 14 ± 1 in SRFA-22, SRFA-6 and SRHA-6 respectively), there were differences in terms of the assembly of the aggregates. These changes in aggregate assemblies can be seen by comparing the hydrophilic and hydrophobic surfaces in the absence and presence of Cu^{2+}, where the introduction of a Cu^{2+} ion resulted in the following changes in surface area: for SRFA-22, there was a 20 % decrease in the hydrophilic surface and 50 % increase in the hydrophobic surface; for SRFA-6, both hydrophilic and hydrophobic surface areas decreased by 20 % and 17 % respectively; and for SRHA-6, the hydrophobic surfaces decreased by 21 % while the hydrophilic surface area remained the same. Overall, it emerged that the addition of a Cu^{2+} ion resulted in the assembly of more compact aggregates with predominantly hydrophobic sites on the surface.

Table 2. Comparison of the characteristics of the modelled systems before and after interaction with organic xenobiotics

System	Without contaminant	Phenol	Toluene	Benzene	
22 FA fragments	21 ± 2	22 ± 2	21 ± 2	17 ± 2	Number of hydrogen bonds
		188	102	−22	ΔE^{A}, kJ mol^{-1}
6 FA fragments	16 ± 2	17 ± 2	16 ± 2	15 ± 1	Number of hydrogen bonds
		158	53	−192	ΔE, kJ mol^{-1}
6 HA fragments	12 ± 1	13 ± 1	11 ± 1	9 ± 1	Number of hydrogen bonds
		134	−46	−214	ΔE, kJ mol^{-1}

[A]ΔE is the difference between the average potential energies of the system modelled with contaminant and without contaminant. To ensure that the energy difference is related only to system conformation and not to the presence of new molecules (contaminants), the molecule of xenobiotic was removed from the system after its optimisation and a single point calculation was run to obtain the potential energy of the system after the contaminant was removed. Thus, the way the aggregate alters to accommodate the contaminant is probed, not just the binding of that contaminant to a pre-existing 'site'.

In terms of potential energy, the addition of the Cu^{2+} ion resulted in a decrease of ~126 kJ mol^{-1} for SFRA-22 and ~209 kJ mol^{-1} for SFRA-6 and SRHA-6. The larger decrease for the systems with large molecular entities was proposed to be related to a more efficient packing owing to the reduction of electrostatic repulsion between negatively charged polar oxygen groups. Similarly, Fomin et al. (2017) reported that the heat of formation of fulvic and humic acids/metal ion complexes differed based on the compound and functional group involved in complex formation. Overall, thermodynamically preferable compact configurations emerged with the introduction of Cu^{2+} into the three HS systems, in agreement with experimental and other computational results (Alvarez-Puebla et al. 2004; Bertoli et al. 2016; Fomin et al. 2017). It should be noted that in this work, Cu-HS interactions were investigated at a very basic level, with the aim of demonstrating the suitability of the ReaxFF to model HS-metal interactions. A detailed analysis of Cu^{2+} and Mg^{2+} interactions with HS in aqueous solution under different conditions will be discussed in a future paper.

The charge redistribution between molecules and atoms that occurred in all three systems and that was related to the addition of the Cu^{2+} ion (for more details refer to Supplementary Material, Text S3, Table S4) allows one to assume that electron shuttling, an emergent property, within HS and NOM in general is, in all likelihood, a very complex process and not the simple transfer of electrons through a conjugated system or stacked aromatic moieties. This point requires further study.

Interactions with organic xenobiotics

Aggregation in a vacuum

A mechanistic understanding of the interactions between organic xenobiotics and HS is important for an understanding of the environmental chemistry of organic contaminants, especially their sequestration. The aim of the modelling in this work was to identify different interaction mechanisms for related organic contaminants with the SRFA and SRHA models by evaluating different characteristics of the models and corelating them with the emergent functioning of HS. Benzene, toluene and phenol were chosen as contaminants. Although these molecules are close in structure they have different steric and polarity properties. Consequently, the interaction mechanism of each with HS should be different (Aristilde and Sposito 2010; Kubicki and Apitz 1999; Saparpakorn et al. 2007) and distinctively affect the conformations of the models. Computational modelling of the three models in a vacuum was correlated with the HS behaviour in a dry state. Subsequent dissolution of the aggregates in water compared the behaviour to the aqueous solution environment and the wetting process of HS.

One molecule of each pollutant was added to each HS model reported above to give nine distinct models. Comparison of the results after simulation with each system at 300 K for only 1 ns did not reveal significant information and might indicate that the simulation time was inadequate. However, a comparison of the potential energies of the systems after their simulation with the temperature raised to 600 K (to overcome interactions with high activation energy and speed up the other interactions), followed by cooling and equilibrating at 300 K revealed certain patterns. The potential energy of the models (Table 2) decreased with the addition of the low polarity compound, benzene. In SRFA models, the benzene molecule was buried, almost perpendicular to the plane of the surface, in a void at the surface. The largest potential energy drop of 214 kJ mol^{-1} was observed for SRHA-6, with the benzene partially in a π-stacking cluster (Fig. 1a). The lower energy decrease that was obtained in other works that modeled similar interactions (Gros et al. 2017; Kubicki and Apitz 1999) may arise from the difference of molecular models that were used, as well as to the use of different computational methods. Although the addition of a benzene molecule resulted in a decrease in the number of hydrogen bonds in each of the three studied systems by 4, 1 and 3 hydrogen bonds for SRFA-22, SRFA-6 and SRHA-6 respectively, the observed formation of larger π-stacking clusters (Fig. 1a) could be the reason for the energy decrease and formation of more thermodynamically stable and preferable configurations. The observed changes in the models is in agreement with experiment, as Nanny and Maza (2001) observed strong sorption of benzene with HA that results in an increase of the hydrophobicity of HA owing to the formation of HA aggregates.

The addition of toluene results in the model pollutant penetrating the surface of all the models in a similar manner to benzene. The changes in potential energy of the models induced by the introduction of toluene (Table 2) imply that SRHA-6 is more amenable to hosting toluene (the energy decreased by 46 kcal mol^{-1}) while the SRFA models are less amenable to hosting toluene (or benzene). This hosting amenability can be explained by the fact that the SRHA-6 system is the least hydrophilic of the systems and also has large hydrophobic domains within a large number of its molecular entities. SFRA-22, however, is the least hydrophobic in nature and does not have molecular entities with large hydrophobic regions, with SRFA-6

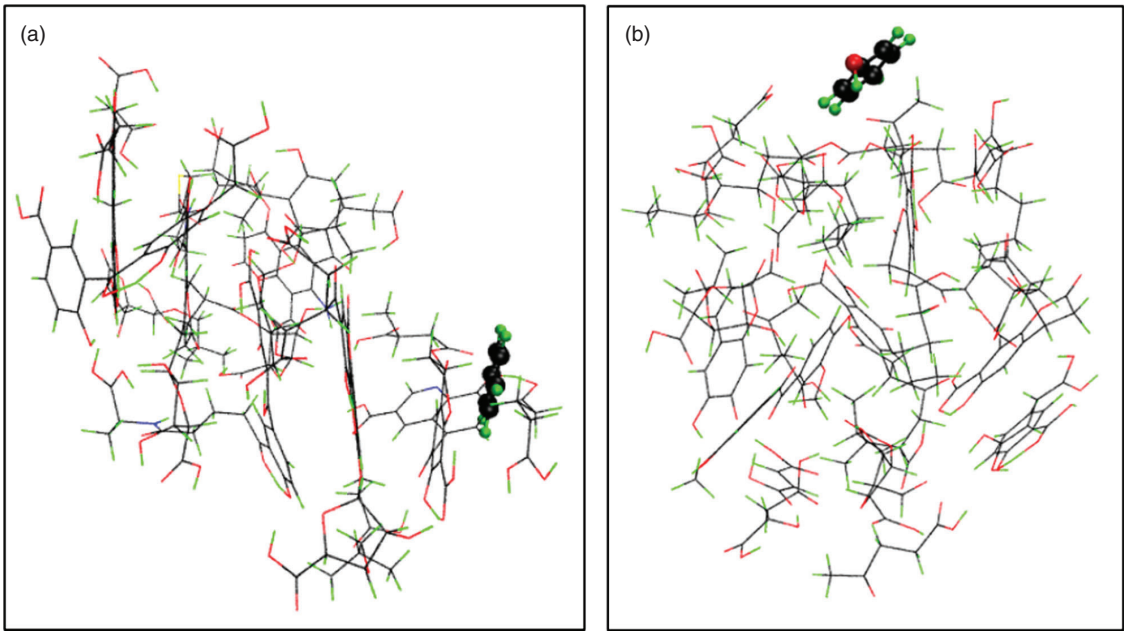

Fig. 1. Snapshots of the structural organisation of the aggregates formed with addition of different pollutants to HS models: (a) SRHA-6 model with addition of benzene, which promotes formation of a π-stacking cluster involving aromatic rings of five molecular fragments; (b) SRFA-22 model with addition of phenol, where phenol is sitting outside of the aggregate and attached to it by an H-bond.

being in the middle. By looking at the difference in potential energy for the benzene versus toluene perturbed systems (Table 2), it can be seen that the hosting of toluene was less favourable. In each model, toluene hosting induces a less favourable interaction within the systems than does benzene. This can be explained by the fact that to host the toluene, its CH_3 groups will be buried into the system assemblage and not stick out. This means that energy must be expended to accommodate the CH_3 group for toluene (less favourable interactions within the system) which is not the case for benzene. The slightly lower energy cost for the SRFA-22 (124 kJ mol^{-1}) can be explained by the ability of this model to create a more diverse set of assemblies and hence its ability to offset the energy cost of hosting the CH_3 group of toluene. We could also conclude that introducing toluene to the systems promotes segregation of hydrophobic and hydrophilic centres. Although toluene has an aromatic ring, its overall structure is not favourable to form π-stacking clusters in comparison with benzene owing to steric hindrance caused by the CH_3 group.

Phenol was the final model pollutant introduced to the system. While from benzene to toluene an increase in steric hinderance was added, owing to the CH_3 group, in going from benzene or toluene to phenol a polar OH group was introduced. This polar group would be expected to interact with the systems under study, mainly through hydrogen bonding, which was also observed in the simulations (Fig. 1b) by increasing the number of H-bonds in each system by 1. However, as the calculations were carried out in a vacuum, the phenol molecule will encounter the hydrophobic surfaces of all three systems. This would lead one to expect the system to have higher potential energies to host phenol versus the other two model pollutants, and this is, indeed, the case (Table 2). The increase in energy (Table 2), especially in SRFA-22, was associated with structural rearrangements, where a larger number of hydrophilic sites turned towards the surface causing an increase of the hydrophilic surface area by 20 % and a decrease of the hydrophobic surface area by 23 %.

Overall, the introduction of the contaminants to the system revealed the emergent functioning of the HS, in particular: rebalancing between different non-covalent interactions, including hydrogen bonding, hydrophobic and π-stacking interactions, and consequently structural rearrangement of the system to host the introduced contaminant at a lower energy cost.

Aggregate dissolution

We turn now to a discussion of MD simulations of the dissolution process after the addition of water to the aggregates formed in a vacuum. The addition of water caused the unfolding of compact aggregates, which resulted in a close to total disaggregation of organic molecules in SRFA-22 (Fig. 2a), as evidenced by the almost tripling of the hydrophilic surface area (Fig. 3b), minor changes in the hydrophobic suface area and a decrease in the number of H-bonds by 9–14 in SRFA-22 (Table S5, Supplementary Material). The observed process was accompanied by an increase of the potential energy calculated for the organic part of SRFA-22 with 1000 water molecules by ~1841 kJ mol^{-1} (Table S6, Supplementary Material) and a drop of the potential energy calculated for the whole system, which included water, by 2.52 % (or 32 702 kJ mol^{-1}). The changes in potential energy confirm the replacement of inter-aggregate H-bonding with organic matter-water surface H-bonds, which is consistent with other computational works (Sutton et al. 2005).

The addition of water to the SRHA-6 model caused the rearrangement of the molecular fragments (Fig. 2b) so that the hydrophilic sites were at the surface (Fig. 3d), which allowed for their interaction with the water molecules, while the hydrophobic components migrated to the centre of the aggregate to

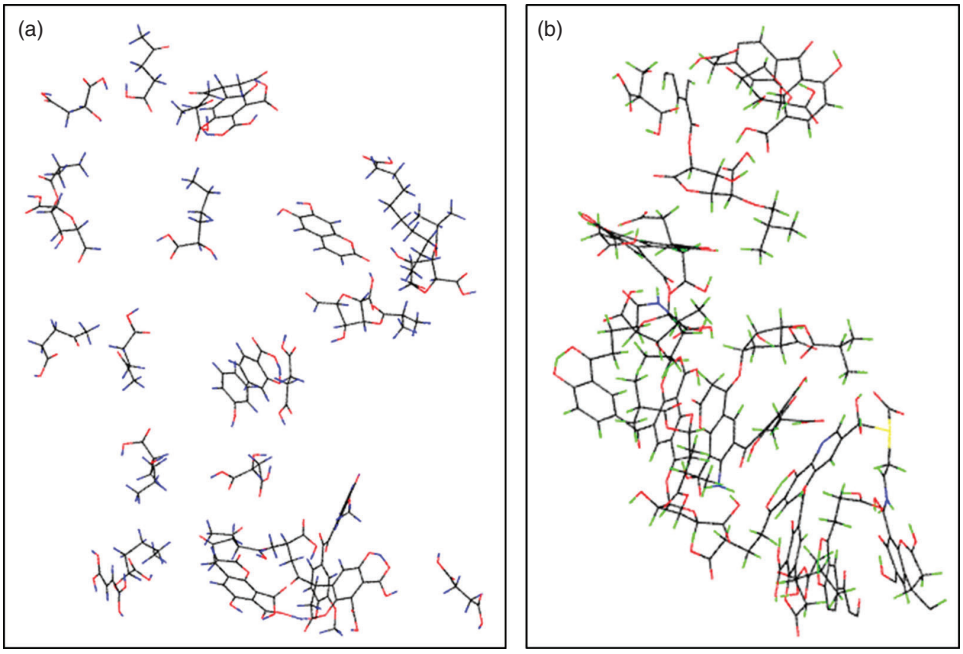

Fig. 2. Snapshots of MD simulations that image a distribution of molecular fragments of HS models in aqueous solution after the dissolution process: (a) SRFA-22 model, in which the aggregate is almost totally dissolved; (b) SRHA-6 model, where molecular fragments hold together preserving the aggregate.

minimise their interaction with the water phase. This rearrangement resulted in a reduction of the inter-aggregate H-bonds by 4, a reduction in the hydrophobic surface area by ~10.5 % (Table S7, Supplementary Material) and an increase in the hydrophilic surface area by ~74 %. Therefore, SRHA-6 was shown to solvate to a lesser extent than SRFA-22 (Fig. 3), as illustrated by the increase in total surface area by ~25 % (versus ~98 % for SRFA-22) and the corresponding increase in the potential energy of the organic fraction by ~732 kJ mol^{-1} versus ~1841 kJ mol^{-1} for SRFA-22.

Comparing the conformations of the SRFA-6, SRHA-6 and SRFA-22 aggregates before and after MD simulations of the dissolution process, we observed that molecules in the aggregates formed from large size fragments, e.g. SRFA-6 and SRHA-6 remained bound together, forming one large cluster (Fig. 2b and Fig. 3c, d), whereas molecular fragments in SRFA-22 were mainly dispersed in the aqueous phase (Fig. 2a and Fig. 3a, b).

The observed difference in dissolution process between SRFA and SRHA models could be related to the difference in experimental methods of extracting fulvic and humic acids (Malcolm et al. 1994; Serkiz and Perdue 1990). FA are expected to dissolve more readily than HA, which is in agreement with the observed results.

Two alternate routes can be proposed for the dissolution process. For the low molecular size FA, in the dry state, one starts off with a compact aggregate of small and/or highly hydrophilic fragments that are H-bonded. When water is added, H-bonding with the water starts to replace the hydrogen H-bonds between the fragments, which results in the separation of the fragments, swelling and the subsequent dissolution of the FA aggregate or assembly. For the large molecular size HA, it is predominantly π-stacking and hydrophobic-induced interactions that are at play in the formation of the dry state aggregates, so that the hydrophobic fragments are exposed at the surface, while the hydrophilic fragments are hidden inside. Upon the addition of water, water molecules penetrate the aggregate and cause it to swell so that there is a rearrangement of the hydrophobic and hydrophilic sites. The rearrangement results in bringing the hydrophilic sites to the surface and to interfacing through H-bonding with the water at the surface. At the same time, the hydrophobic core holds the assembly together.

The addition of various contaminants also affected the dissolution process of organic matter. The addition of hydrophobic pollutants, benzene or toluene, resulted in a lower dissolution rate of organic matter. Thus, after running MD simulations for 2.3 ns, the largest total surface area was observed in models with phenol (Table S7, Supplementary Material), whereas aggregates with benzene and toluene remained more compact. Similarly, the largest decrease in the number of H-bonds upon dissolution (Table S5, Supplementary Material) was observed in the models with phenol. The most significant drop in H-bonding, by 17, occurred in SRFA-22 with phenol, where H-bonding is the main aggregation mechanism. As it was shown above, the addition of a hydrophobic contaminant in dry conditions promotes the formation of more thermodynamically stable aggregates of organic matter that eventually become less soluble. Thus, the increase of the hydrophilic surface area for the models with the addition of benzene after dissolution was less compared with the models with phenol, and it was twice as low in SRHA-6. Although the energy of the whole systems including water decreased, the potential energy calculated for the organic fraction after dissolution increased (Table S6, Supplementary Material). The highest increase of ~1243 kJ mol^{-1} (in average) occurred in models containing phenol compared with the replicas with benzene (~1046 kJ mol^{-1}) or toluene (~1184 kJ mol^{-1}).

Our results imply that phenol interacting through H-bonding does not strongly affect the dissolution of HS compared with the more hydrophobic toluene and benzene, which decrease the dissolution rate by holding molecular fragments together through hydrophobic and π-stacking interactions. A similar

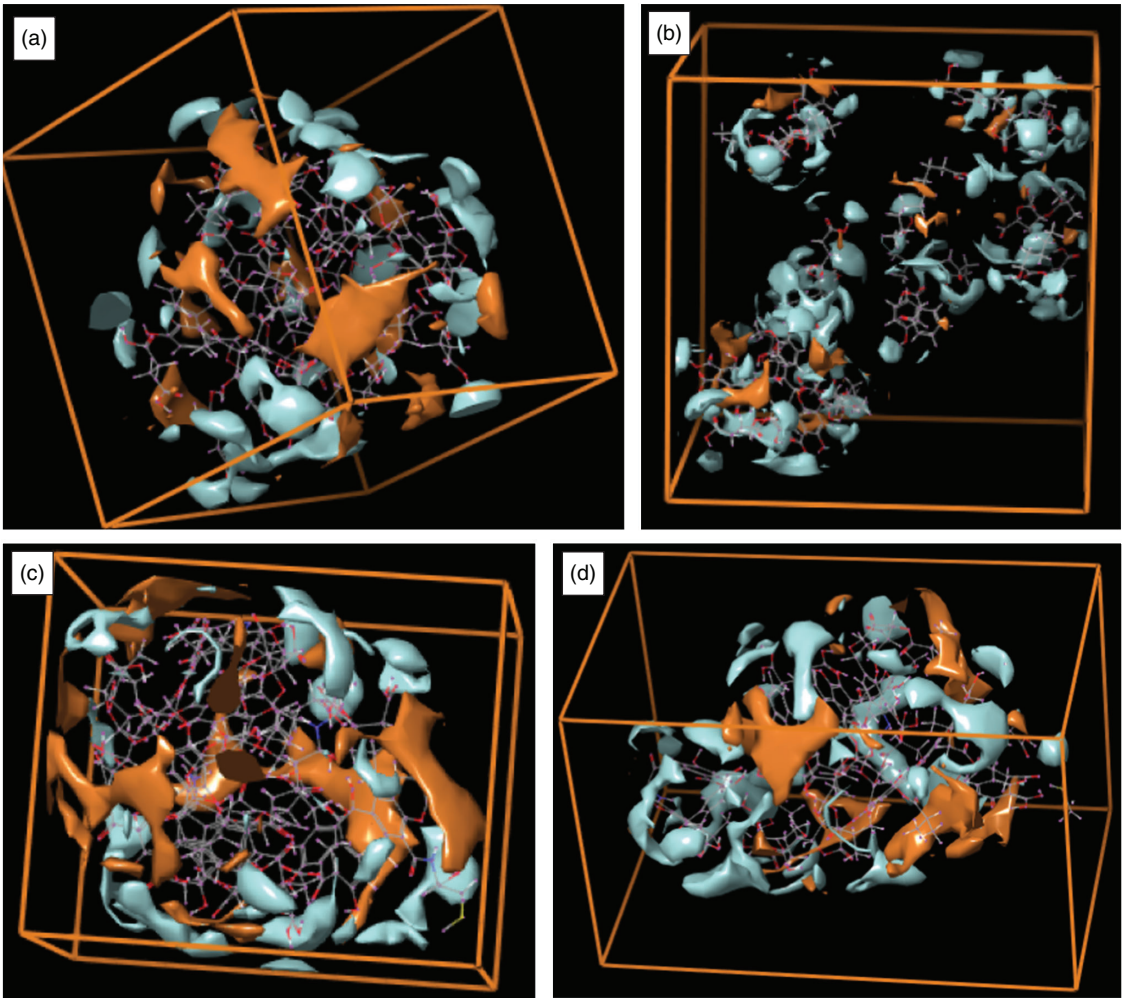

Fig. 3. Distribution of hydrophilic (blue color) and hydrophobic (orange color) surface areas in HS models before and after dissolution process. (a) SRFA-22 model with phenol before dissolution with the presence of large hydrophobic sites at the surface; (b) SRFA-22 model with phenol after dissolution, where the hydrophilic surface is predominant; (c, d) SRHA-6 model with benzene before and after dissolution respectively, with a slight increase of the hydrophilic surface area after dissolution. Each picture corresponds to a single snapshot from an MD simulation.

conclusion was made by Ahmed et al. (2016) who studied the interactions of soil organic matter with contaminants using the COSMO method. These observations by computational methods are consistent with the NMR experimental results on the interaction of monoaromatic compounds with HA (Nanny 1999).

Conclusions

The present study provides important insight into the emergent behaviour of complex organic mixtures using, as an example, HS, and relates emergent functioning to system composition and molecular size. The obtained MD results of HS computational modelling stress the importance of non-covalent interactions, which include H-bonding, hydrophilic/hydrophobic, π-stacking and so forth, for the emergent behaviour of the systems, such as aggregation/dissolution and interaction with polar and non-polar pollutants. Thus, the following outstanding observations emerge: (1) in organic matter composed mainly from small hydrophilic molecular moieties, H-bonding is the dominant force that drives the aggregation process in dry conditions. However, in wet conditions, the H-bonding interactions are weak and result in quick dissolution of the aggregates; (2) In a dry environment, the structural configuration of the aggregates of the SRFA-6 model, which contains large molecular moieties with aromatic sites, though still predominantly hydrophilic molecules, is the result of a competition between H-bonding and hydrophobic/π-stacking interactions. The subsequent addition of water to the SRFA-6 model causes an increase of the hydrophilic surface areas of the aggregates, though the hydrophobic centres formed prevent complete dissolution of the aggregates; (3) In the hydrophobic, highly aromatic system, SRHA-6, with large molecular size moieties, the hydrophobic-induced and π-stacking interactions govern the aggregation process and the structural configuration of the formed labile assemblies. Consequently, the wetting process requires a much longer time and causes the rearrangement of the clusters in a way that allows hydrophobic sites to move inside, which holds the molecules together, and the hydrophilic sites turn to the outside, which interacts with the aqueous interface.

The picture that arises from this study could have implications for several known phenomena for HS, such as sorption and sequestration of organic contaminants and solute transport. In particular, it was shown that polar phenol interacts with HS

through a much weaker H-bonding mechanism compared with the hydrophobic benzene, which mainly interacts through hydrophobic-induced and π-stacking mechanisms, and results in a much stronger sorption by HA. The reason for the strong interactions between hydrophobic pollutants, e.g. benzene, and highly aromatic HA is the emergent structural rearrangement of the labile assemblies with the creation of space inside the aggregate to accommodate the contaminant. The resultant structural changes lead to the formation of more stable and more thermodynamically preferred systems.

Computer simulations could be a complementary tool to experiments to improve our understanding of the interaction mechanisms between HS and organic pollutants and their complex dynamics at the molecular level. Therefore, the results of this work could have implications for environmental science in several respects: studying the mechanism of sequestration of polar, non-polar, aromatic and aliphatic pollutants in soil to determine the mechanism of possible transformations; understanding the mechanisms of sorption and desorption of different contaminants in soil and water, and consequently, their toxicity to the surrounding organisms; investigation of sorption behaviour, including sorption hysteresis (Schaumann and Thiele-Bruhn 2011), 'irreversible' sorption and the conditioning effect (enhanced repeat sorption) (Sander et al. 2006).

Although computational models can predict certain interaction processes of HS, the predicted interaction energies and majority of the properties of the models cannot be verified directly in experimental studies. The modelling rather gives insights into the potential of HS to react and interact with its environment. However, it is still possible to make a comparison and approximations. Thus, the models interact differently with the studied contaminants, since different mechanisms are involved in the aggregation process. Another important conclusion that could be made is that the added contaminant also affects the functional behaviour and structural configuration of the system. Thus HS, treated as labile aggregates, could undergo structural rearrangements to create space inside the aggregate to accommodate the contaminant if it causes the formation of a more stable and more thermodynamically preferable system. Therefore, two possible mechanisms of contaminant adsorption could be involved: surface adsorption and absorption of pollutant molecules into internal voids in the HS structure with structural rearrangement of the aggregates.

Overall, the combination of dynamic combinatorial chemistry, agent-based modelling and supramolecular labile aggregation approaches allows the realistic modelling of the emergent functional behaviour of complex organic mixtures, such as HS. Thus, the application of the ReaxFF force field allows one step forward in modelling interactions of large systems, which include proton transfer.

Supplementary material

Supplementary material and data in support of this paper are available on the Journal's website.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors are grateful to NSERC-Canada for ongoing Discovery Grants. They also gratefully acknowledge the provision of computational resources by Compute Canada/WestGrid.

References

Ahmed AA, Thiele-Bruhn S, Leinweber P, Kühn O (2016). Towards a molecular level understanding of the sulfanilamide-soil organic matter-interaction. *The Science of the Total Environment* **559**, 347–355. doi:10.1016/J.SCITOTENV.2016.03.136

Alvarez-Puebla RA, Valenzuela-Calahorro C, Garrido JJ (2004). Retention of Co(II), Ni(II), and Cu(II) on a Purified Brown Humic Acid. Modeling and Characterization of the Sorption Process. *Langmuir* **20**, 3657–3664. doi:10.1021/LA0363231

Alvarez-Puebla RA, Valenzuela-Calahorro C, Garrido JJ (2006). Theoretical study on fulvic acid structure, conformation and aggregation. *The Science of the Total Environment* **358**, 243–254. doi:10.1016/J.SCITOTENV.2004.11.026

Aristilde L, Sposito G (2010). Binding of ciprofloxacin by humic substances: A molecular dynamics study. *Environmental Toxicology and Chemistry* **29**, 90–98. doi:10.1002/ETC.19

Ashraf C, Jain A, Xuan Y, van Duin ACT (2017). ReaxFF based molecular dynamics simulations of ignition front propagation in hydrocarbon/oxygen mixtures under high temperature and pressure conditions. *Physical Chemistry Chemical Physics* **19**, 5004–5017. doi:10.1039/C6CP08164A

Bar-Yam Y (2002). General features of complex systems. Available at https://www.researchgate.net/publication/246294756_General_Features_of_Complex_Systems [verified 11 September 2017]

Bertoli AC, Garcia JS, Trevisan MG, Ramalho TC, Freitas MP (2016). Interactions fulvate-metal (Zn2+, Cu2+ and Fe2+): theoretical investigation of thermodynamic, structural and spectroscopic properties. *Biometals* **29**, 275–285. doi:10.1007/S10534-016-9914-8

Bruccoleri RE, Karplus M (1990). Conformational sampling using high-temperature molecular dynamics. *Biopolymers* **29**, 1847–1862. doi:10.1002/BIP.360291415

Bruccoleri AG, Sorenson BT, Langford CH (2001). Molecular modelling of humic structures. In 'Humic substances: structures, models and functions'. (Eds EA Ghabbour, G Davies) pp. 193–208. (Royal Society of Chemistry: London)

Cabaniss SE (2011). Forward Modeling of Metal Complexation by NOM: II. Prediction of Binding Site Properties. *Environmental Science & Technology* **45**, 3202–3209. doi:10.1021/ES102408W

Cabaniss SE, Madey G, Leff L, Maurice PA, Wetzel R (2005). A Stochastic Model for the Synthesis and Degradation of Natural Organic Matter. Part I. Data Structures and Reaction Kinetics. *Biogeochemistry* **76**, 319–347. doi:10.1007/S10533-005-6895-Z

Cook RL, Langford CH (1998). Structural characterization of a fulvic acid and a humic acid using solid-state ramp-CP-MAS 13C nuclear magnetic resonance. *Environmental Science & Technology* **32**, 719–725. doi:10.1021/ES970488C

Cook RL, Langford CH (1999*a*). Ramped amplitude cross polarisation magic angle spinning NMR (Ramp-CP-MAS-NMR): A technique for quantitative study of the composition of solid state polymers. *Polymer News* **24**, 6–15.

Cook RL, Langford CH (1999*b*). A biogeopolymeric view of humic substances with application to paramagnetic metal effects on ^{13}C NMR. In 'Understanding humic substances'. (Eds EA Ghabbour, G Davies) pp. 31–48. (Woodhead Publishing: Cambridge, UK)

Corbett PT, Leclaire J, Vial L, West KR, Wietor JL, Sanders JKM, Otto S (2006). Dynamic Combinatorial Chemistry. *Chemical Reviews* **106**, 3652–3711. doi:10.1021/CR020452P

Drosos M, Nebbioso A, Mazzei P, Vinci G, Spaccini R, Piccolo A (2017). A molecular zoom into soil Humeome by a direct sequential chemical fractionation of soil. *The Science of the Total Environment* **586**, 807–816. doi:10.1016/J.SCITOTENV.2017.02.059

Fomin VN, Gogol DB, Rozhkovoy IE, Ponomarev DL (2017). Quantum chemical and thermodynamic calculations of fulvic and humic copper complexes in reactions of malachite and azurite formation. *Applied Geochemistry* **79**, 9–16. doi:10.1016/J.APGEOCHEM.2017.02.002

Gros P, Ahmed A, Kühn O, Leinweber P (2017). Glyphosate binding in soil as revealed by sorption experiments and quantum-chemical modeling. *The Science of the Total Environment* **586**, 527–535. doi:10.1016/J.SCITOTENV.2017.02.007

Gumbart JC (2007). HBonds plugin, version 1.2. Available at http://www.ks.uiuc.edu/Research/vmd/plugins/hbonds/ [verified 12 November 2017]

Hertkorn N, Frommberger M, Witt M, Koch BP, Schmitt-Kopplin P, Perdue EM (2008). Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Analytical Chemistry* **80**, 8908–8919. doi:10.1021/AC800464G

Huang Y, Xiang X, Madey G, Cabaniss SE (2005). Agent-based scientific simulation. *Computing in Science & Engineering* **7**, 22–29. doi:10.1109/MCSE.2005.7

Humphrey W, Dalke A, Schulten K (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38. doi:10.1016/0263-7855(96)00018-5

Hunter AD (1997). ACD/ChemSketch 1.0 (freeware); ACD/ChemSketch 2.0 and its Tautomers, Dictionary, and 3D Plug-ins; ACD/HNMR 2.0; ACD/CNMR 2.0. *Journal of Chemical Education* **74**, 905–906. doi:10.1021/ED074P905

International Humic Substances Society (IHSS) (2019). Available at http://humic-substances.org/ [verified 12 November 2017]

Jansen SA, Malaty M, Nwabara S, Johnson E, Ghabbour E, Davies G, Varnum JM (1996). Structural modeling in humic acids. *Materials Science and Engineering C* **4**, 175–179. doi:10.1016/S0928-4931(96)00151-8

Kubicki JD, Apitz SE (1999). Models of natural organic matter and interactions with organic contaminants. *Organic Geochemistry* **30**, 911–927. doi:10.1016/S0146-6380(99)00075-3

Leenheer JA, McKnight DM, Thurman EM, MacCarthy P (1994). Structural components and proposed structural models of fulvic acid from the Suwannee River. In 'Humic substances in the Suwannee River, Georgia: interactions, properties, and proposed structures', US Geological Survey, No 87–557 (Eds RC Averett, JA Leenheer, DM McKnight, KA Thorn) pp. 195–212. (US G.P.O., US Geological Survey, Map Distribution: Georgia)

Litvina M, Todoruk TR, Langford CH (2003). Composition and Structure of Agents Responsible for Development of Water Repellency in Soils following Oil Contamination. *Environmental Science & Technology* **37**, 2883–2888. doi:10.1021/ES026296L

Malcolm RL, Aiken GR, Bowles EC, Malcolm JD (1994). Isolation of fulvic and humic acids from the Suwannee River. In 'Humic substances in the Suwannee River, Georgia: interactions, properties, and proposed structures', US Geological Survey, No 87–557 (Eds RC Averett, JA Leenheer, DM McKnight, KA Thorn) pp. 195–212. (US G.P.O., US Geological Survey, Map Distribution: Georgia)

Martínez L, Andrade R, Birgin EG, Martínez JM (2009). PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **30**, 2157–2164. doi:10.1002/JCC.21224

Monti S, Corozzi A, Fristrup P, Joshi KL, Shin YK, Oelschlaeger P, van Duin ACT, Barone V (2013). Exploring the conformational and reactive dynamics of biomolecules in solution using an extended version of the glycine reactive force field. *Physical Chemistry Chemical Physics* **15**, 15062–15077. doi:10.1039/C3CP51931G

Nanny MA (1999). Deuterium NMR characterization of noncovalent interactions between monoaromatic compounds and fulvic acids. *Organic Geochemistry* **30**, 901–909. doi:10.1016/S0146-6380(99)00074-1

Nanny MA, Maza JP (2001). Noncovalent Interactions between Monoaromatic Compounds and Dissolved Humic Acids: A Deuterium NMR T1 Relaxation Study. *Environmental Science & Technology* **35**, 379–384. doi:10.1021/ES0012927

Nebbioso A, Piccolo A (2012). Advances in humeomics: Enhanced structural identification of humic molecules after size fractionation of a soil humic acid. *Analytica Chimica Acta* **720**, 77–90. doi:10.1016/J.ACA.2012.01.027

Orsi M (2014). Molecular dynamics simulation of humic substances. *Chemical and Biological Technologies in Agriculture* **1**, 10. doi:10.1186/S40538-014-0010-4

Perminova IV, Hatfield K (2005). Remediation chemistry of humic substances: theory and implications for technology. In 'Use of humic substances to remediate polluted environments: from theory to practice. IV. Earth and environmental sciences'. (Eds IV Perminova, K Hatfield, N Hertkorn) pp. 3–37. (Springer: Dordrecht, the Netherlands)

Petrov D, Tunega D, Gerzabek MH, Oostenbrink C (2017). Molecular Dynamics Simulations of the Standard Leonardite Humic Acid: Microscopic Analysis of the Structure and Dynamics. *Environmental Science & Technology* **51**, 5414–5424. doi:10.1021/ACS.EST.7B00266

Piccolo A (2002). The supramolecular structure of humic substances: A novel understanding of humus chemistry and implications in soil science. *Advances in Agronomy* **75**, 57–134. doi:10.1016/S0065-2113(02)75003-7

Rahaman O, van Duin ACT, Goddard WA, Doren DJ (2011). Development of a ReaxFF Reactive Force Field for Glycine and Application to Solvent Effect and Tautomerization. *The Journal of Physical Chemistry B* **115**, 249–261. doi:10.1021/JP108642R

Reemtsma T (2009). Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry. *Journal of Chromatography A* **1216**, 3687–3701. doi:10.1016/J.CHROMA.2009.02.033

Reemtsma T, These A, Springer A, Linscheid M (2008). Differences in the molecular composition of fulvic acid size fractions detected by size-exclusion chromatography–on line Fourier transform ion cyclotron resonance (FTICR–) mass spectrometry. *Water Research* **42**, 63–72. doi:10.1016/J.WATRES.2007.06.063

Remucal CK, Cory RM, Sander M, McNeill K (2012). Low Molecular Weight Components in an Aquatic Humic Substance As Characterized by Membrane Dialysis and Orbitrap Mass Spectrometry. *Environmental Science & Technology* **46**, 9350–9359. doi:10.1021/ES302468Q

Sander M, Lu Y, Pignatello JJ (2006). Conditioning-annealing studies of natural organic matter solids linking irreversible sorption to irreversible structural expansion. *Environmental Science & Technology* **40**, 170–178. doi:10.1021/ES0506253

Saparpakorn P, Kim JH, Hannongbua S (2007). Investigation on the Binding of Polycyclic AromaticHydrocarbons with Soil Organic Matter: A Theoretical Approach. *Molecules* **12**, 703–715. doi:10.3390/12040703

Schaumann GE, Thiele-Bruhn S (2011). Molecular modeling of soil organic matter: Squaring the circle?. *Geoderma* **166**, 1–14. doi:10.1016/J.GEODERMA.2011.04.024

Schrodinger Software (2015). Maestro 10.2 user manual (Schrodinger Press: New York, NY). Available at http://gohom.win/ManualHom/Schrodinger/Schrodinger_2015-2_docs/maestro/maestro_user_manual.pdf [verified 15 December 2017]

Schulten HR, Leinweber P (2000). New insights into organic-mineral particles: composition, properties and models of molecular structure. *Biology and Fertility of Soils* **30**, 399–432. doi:10.1007/S003740050020

Schulten HR, Thomsen M, Carlsen L (2001). Humic complexes of diethyl phthalate: molecular modelling of the sorption process. *Chemosphere* **45**, 357–369. doi:10.1016/S0045-6535(00)00590-7

Sein LT, Varnum JM, Jansen SA (1999). Conformational Modeling of a New Building Block of Humic Acid: Approaches to the Lowest Energy Conformer. *Environmental Science & Technology* **33**, 546–552. doi:10.1021/ES9805324

Serkiz SM, Perdue EM (1990). Isolation of Dissolved Organic Matter from the Suwannee River Using Reverse Osmosis. *Water Research* **24**, 911–916. doi:10.1016/0043-1354(90)90142-S

Stenson AC, Marshall AG, Cooper WT (2003). Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Analytical Chemistry* **75**, 1275–1284. doi:10.1021/AC026106P

Sündermann A, Solc R, Tunega D, Haberhauer G, Gerzabek MH, Oostenbrink C (2015). Vienna Soil-Organic-Matter Modeler—Generating condensed-phase models of humic substances. *Journal of Molecular Graphics & Modelling* **62**, 253–261. doi:10.1016/J.JMGM.2015.10.007

Sutton R, Sposito G (2005). Molecular Structure in Soil Humic Substances: The New View. *Environmental Science & Technology* **39**, 9009–9015. doi:10.1021/ES050778Q

Sutton R, Sposito G, Diallo MS, Schulten HR (2005). Molecular simulation of a model of dissolved organic matter. *Environmental Toxicology and Chemistry* **24**, 1902–1911. doi:10.1897/04-567R.1

Tan KH (2014). 'Humic matter in soil and the environment: principles and controversies, 2nd edn.' (CRC Press: Boca Raton, FL)

Underdown AW, Langford CH, Gamble DS (1981). Light scattering of a polydisperse fulvic acid. *Analytical Chemistry* **53**, 2139–2140. doi:10.1021/AC00236A046

van Duin ACT, Dasgupta S, Lorant F, Goddard WA (2001). ReaxFF: A Reactive Force Field for Hydrocarbons. *The Journal of Physical Chemistry A* **105**, 9396–9409. doi:10.1021/JP004368U

Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 671–690. doi:10.1002/JCC.21367

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry* **25**, 1157–1174. doi:10.1002/JCC.20035

Wershaw RL (1986). A new model for humic materials and their interactions with hydrophobic organic chemicals in soil-water or sediment-water systems. *Journal of Contaminant Hydrology* **1**, 29–45. doi:10.1016/0169-7722(86)90005-7

Xue H, Sigg L (1999). Comparison of the complexation of Cu and Cd by humic or fulvic acids and by ligands observed in lake waters. *Aquatic Geochemistry* **5**, 313–335. doi:10.1023/A:1009679819002

Handling Editor: Kevin Wilkinson